COLLOQUIA

*Revista de Pensamiento y Cultura Vol. **12** (2025), págs. 535 - 557. ISSN: 1390-8731*
DOI 10.31207/colloquia.v12i1.197

La Inteligencia Artificial en la formulación y personalización de políticas públicas: Entre la eficiencia, la equidad y el riesgo democrático

Artificial Intelligence in the formulation and customization of public policies: Between efficiency, equity, and democratic risk

Dayanara Noelly Solano Carpio
dnsolano@utpl.edu.ec
Universidad Técnica Particular de Loja
ORCID ID 0009-0003-4871-1530

Boris Raúl Ochoa Ordóñez
brochoa@utpl.edu.ec
Universidad Técnica Particular de Loja
ORCID ID 0000-0003-4928-0915

Resumen: La incorporación de sistemas de inteligencia artificial (IA) en el ciclo de las políticas públicas está transformando de manera sustantiva la manera en que los Estados diagnostican problemas, diseñan intervenciones, asignan recursos y evalúan resultados. Este artículo examina críticamente cómo la IA reconfigura las relaciones entre eficiencia administrativa, equidad distributiva, transparencia institucional y legitimidad democrática, y propone un marco operativo para su adopción responsable en contextos públicos. Este análisis se sustenta en la pregunta ¿bajo qué condiciones los beneficios productos del uso de la IA justifican los riesgos que representan para la gobernanza democrática contemporánea? Para responder esta pregunta, esta investigación elabora un análisis conceptual y metodológico con base en cuatro pilares: eficiencia, equidad, transparencia y legitimidad. En cada uno de estos se empieza con una definición operativa, pasando por señales verificables que

Recepción: 10-10-2025 Aceptación: 08-12-2025 Publicación electrónica: 20-12-2025

permiten evaluar su nivel de cumplimiento y, por último, definir los límites normativos que deberían trazar el camino de su aplicación. De igual manera, la investigación marca una diferencia entre personalización política y personalización administrativa, evidenciando que la personalización administrativa mejora la concentración y provisión de servicios. No obstante, también puede incrementar las desigualdades si es que existen sesgos en los datos o una baja supervisión humana. El artículo plantea una secuencia de gobernanza antes, durante y después de la implementación que enlaza evaluación de impacto algorítmico, dirección humana en puntos clave, auditoría externa periódica, trazabilidad, publicidad de fundamentos y mecanismos prácticos de apelación. Partiendo de casos internacionales de salud, protección social, educación, se identifican requisitos que autorizan y supuestos de veto para asegurar que la eficiencia no se priorice sobre la justicia social, la no discriminación y la correcta rendición de cuentas. Esta investigación culmina señalando que la IA puede ayudar a fortalecer la capacidad del estado, pero solo si su implementación se realiza en una arquitectura institucional que resguarde la participación ciudadana, el control democrático y la transparencia. La IA no sustituye el juicio humano ni el proceso deliberativo; lo complementa cuando existen salvaguardas claras, estándares verificables y una supervisión continúa orientada al bienestar público.

Palabras clave: Inteligencia artificial; política pública; toma de decisiones algorítmica; gobernanza digital.

Abstract: The integration of artificial intelligence (AI) systems into the public policy cycle is substantially transforming how states diagnose problems, design interventions, allocate resources, and evaluate outcomes. This article critically examines how AI reconfigures the relationships between administrative efficiency, distributive equity, institutional transparency, and democratic legitimacy, and proposes an operational framework for its responsible adoption in public contexts. This analysis is based on the question: under what conditions do the benefits of AI justify the risks it poses to contemporary democratic governance? To answer this question, this research develops a conceptual and methodological analysis based on four pillars: efficiency, equity, transparency, and legitimacy. Each pillar begins with an operational definition, moves on to verifiable indicators that allow for the assessment of its level of compliance, and finally defines the normative limits that should guide its application. Similarly, the research distinguishes between political personalization and administrative personalization, demonstrating that administrative personalization improves the concentration and provision of services.

However, it can also exacerbate inequalities if there are biases in the data or weak human oversight. The article proposes a governance sequence before, during, and after implementation that links algorithmic impact assessment, human guidance at key points, periodic external audits, traceability, disclosure of rationale, and practical appeals mechanisms. Drawing on international cases in health, social protection, and education, it identifies enabling requirements and veto assumptions to ensure that efficiency is not prioritized over social justice, non-discrimination, and proper accountability. This research concludes by noting that AI can help strengthen state capacity, but only if its implementation takes place within an institutional framework that safeguards citizen participation, democratic control, and transparency. AI does not replace human judgment or the deliberative process; it complements them when there are clear safeguards, verifiable standards, and continuous oversight focused on the public good.

Keywords: Artificial intelligence; public policy; algorithmic decision-making; digital governance.

Introducción

La implementación de sistemas de inteligencia artificial (IA) en la gestión públicas se ha convertido actualmente en uno de los focos de atención en el debate académico y político sobre el futuro del Estado, la democracia y la justicia social. Su uso en las diferentes etapas del ciclo de creación, implementación y evaluación de políticas públicas augura incrementar la eficiencia administrativa, optimizar la focalización de recursos y aumentar la capacidad del Estado para prever riesgos y organizar intervenciones. No obstante los beneficios se opacan por temas como desigualdad estructural, sesgos logarítmicos, toma de decisiones automáticas y sin transparencia, dependencia tecnológica y débil control ciudadano.

Estudios recientes están de acuerdo en que estos problemas son parte fundamental de la manera en la que los sistemas algorítmicos reorganizan la

relación entre el Estado, los ciudadanos y los procesos de elección democrática (Eubanks, 2018; Yeung, 2018).

Aún con el creciente interés en la implementación de la IA en los gobiernos nacionales y locales, todavía perduran brechas analíticas y normativas que obstaculizan valorar de forma integral su huella en la capacidad estatal y en la legitimidad institucional. Parte importante de los estudios se enfocan en aplicaciones sectoriales como la salud, educación, protección social o en diagnósticos generales sobre riesgos éticos, sin elaborar marcos operativos que permitan a las administraciones públicas disponer cuándo, cómo y en qué contextos la IA puede utilizarse de forma responsable (Kroll, 2020; Shneiderman, 2022). Esta ausencia de criterios aplicables genera dos riesgos simultáneos: por una parte, procesos de automatización precipitados que profundizan desigualdades; por otra, restricciones excesivas que limitan la innovación pública aun cuando podrían implementarse salvaguardas adecuadas.

Este artículo parte de una pregunta central: ¿en qué condiciones los beneficios asociados al uso de IA justifican los riesgos que introduce para la gobernanza democrática contemporánea? Esta interrogante no pretende confrontar eficiencia y democracia como extremos incompatibles, sino más bien identificar las condiciones en las que la IA puede fortalecer la capacidad estatal sin poner en tela de juicio los valores fundamentales de equidad, transparencia y legitimidad. En consecuencia, se aplica un enfoque conceptual-analítico estructurado en cuatro pilares: eficiencia, equidad, transparencia y legitimidad.

Cada pilar se despliega desde una definición operativa, los mecanismos que explican como la IA afecta dicho ámbito, las señales verificables que permiten evaluar su nivel de cumplimiento, y los límites normativos que deberían ser los ejes sobre su utilización.

La principal contribución de este artículo es convertir un análisis general sobre ventajas y riesgos en un marco evaluativo y operativo aplicable por instituciones públicas. En contraste con las aproximaciones basadas en

listados de riesgos o en manuales de buenas prácticas, este estudio propone un enfoque secuencial que enlaza lo que debe suceder antes, durante y después del despliegue de sistemas algorítmicos. Esto es: (a) la realización de evaluaciones de impacto algorítmico previas a la implementación, (b) establecimiento de umbrales mínimos para la automatización de procesos, (c) la obligatoriedad de supervisión humana en puntos clave, (d) mecanismos fuertes de trazabilidad, auditoría externa y apelación ciudadana una vez que la IA se encuentra funcionando (Kaminski & Malgieri, 2021; Raji et al., 2020).

La introducción conceptualiza la diferencia entre personalización política que tramita procesos de comunicación, liderazgo y representación, y la personalización administrativa que define las formas en las que el Estado modifica los recursos y servicios en función de las características individuales o grupales de la población.

Esta diferenciación no es irrelevante ya que confundirlas deriva en diagnóstico equivocados y sesgos en la interpretación. Este artículo se enfoca exclusivamente en la personalización administrativa y analiza sus efectos sobre la equidad, eficiencia y legitimidad.

Por último, en lo que al alcance del texto tiene relación, este no es un estudio empírico ni busca evaluar un caso nacional o local específico, sino construir un marco conceptual anclado en bibliografía interdisciplinaria reciente (2020 – 2025) y una evidencia comparada sobre la implementación de IA en servicios públicos. Se busca establecer criterios operativos aplicables a instituciones de diversos niveles de gobierno interesadas en incorporar la IA de forma responsable, transparente y acorde a los principios democráticos.

En síntesis la introducción establece la relevancia del debate define el enfoque analítico, presenta la pregunta de investigación y destaca de forma breve la contribución del artículo en el ámbito de la gobernanza algorítmica y las políticas públicas.

Metodología

Esta investigación adopta un enfoque metodológico mixto que combina un análisis conceptual-analítico con una revisión narrativa estructurada de literatura interdisciplinaria reciente. Este diseño metodológico permite avanzar desde una comprensión teórica de los mecanismos mediante los cuales la inteligencia artificial (IA) transforma la acción pública hacia la identificación de patrones, tensiones y criterios aplicables en contextos gubernamentales concretos. La mezcla de ambos enfoques cumple con la necesidad de articular dentro de un mismo marco, los fundamentos normativos de la gobernanza algorítmica y la evidencia empírica reciente sobre el uso de la IA en sectores como la salud, educación, administración pública digital y protección social.

La parte conceptual-analítica empieza con la revisión de teoría relacionada con la regulación algorítmica, equidad, transparencia y legitimidad democrática. Busca definir de manera operativa los conceptos de eficiencia, equidad, transparencia y legitimidad.

Estos conceptos están formados por criterios de validez, indicadores para medir su efectividad y características causales que permiten analizar la forma en que la IA interviene en la dinámica estatal.

Esta investigación tomó como referente literatura especializada en la ética en el uso de la IA, políticas públicas y gobernanza digital, permitiendo la convergencia entre la teoría, los riesgos documentados y evaluaciones institucionales.

El segundo componente de la metodología hace referencia a la revisión narrativa estructurada de literatura publicada entre 2020 y 2025. La delimitación de estos años hace referencia al rápido avance que han tenido las tecnologías de aprendizaje automático y al surgimiento de marcos regulatorios recientes que rediseñan los estándares de gobernanza algorítmica.

Se incluyeron en el análisis, informes técnicos de organismos multilaterales, artículos científicos, documentos de políticas públicas y

estudios comparados y enfocados en el uso de la IA en gobiernos nacionales y locales. Para la revisión bibliográfica se consultó bases de datos académicas como Google Scholar, Scopus, entre otras y se complementó con información encontrada en repositorios especializados en gobernanza digital y políticas públicas.

Se eligió las fuentes de información priorizando los estudios referentes a la aplicación de la IA en funciones del Estado, investigaciones que evalúan el impacto de la IA en los cuatro pilares antes mencionados, documentos que se enfocan en mecanismos de auditoría y control y bibliografía que define la ética en el uso de IA.

No se usaron documentos que eran puramente técnicos ni que no guardaban relación con la política pública. También se excluyeron estudios en los que se analiza el uso de la IA con fines de lucro y, por tema de actualidad, tampoco artículos publicados antes del 2019.

En principio se recolecto información enfocada en los cuatro pilares (eficiencia, equidad, transparencia y legitimidad democrática), y luego se buscó información relacionada con formas de dar a entender los procesos tecnológicos detrás de la IA, maneras de procurar la protección institucional, auditorías y control ciudadano.

Este enfoque mixto permitió diseñar una secuencia operativa de gobernanza algorítmica construida mediante la comparación entre nuevos marcos normativos y evidencia empírica sobre implementaciones exitosas y fallidas. Esta secuencia busca transformar principios abstractos en pasos verificables antes, durante y después del despliegue de sistemas algorítmicos en el sector público.

Finalmente, se reconocen las limitaciones metodológicas del estudio ya que no corresponde a una revisión sistemática exhaustiva ni a estudios de campo, sino más bien a una aproximación híbrida dirigida a la construcción de una marco conceptual robusto y operativo. Aun así, este enfoque permite detectar tendencias consistentes en la literatura reciente, la reconstrucción de

mecanismos de impacto y la elaboración de criterios prácticos para el fortalecimiento de la gobernanza democrática en contextos de rápida digitalización.

Marco teórico y ejes analíticos

El análisis de la inteligencia artificial (IA) en el sector público requiere un marco teórico capaz de articular dimensiones administrativas, tecnológicas y normativas. Este artículo acoge una visión que entiende a la IA como un conjunto de sistemas algorítmicos que transfiguran la forma en que los Estados analizan la información, coordinan intervenciones y reparten los bienes públicos. Empero, estas transfiguraciones no se limitan al ámbito técnico, también reajustan relaciones de poder, patrones de equidad y la misma estructura de la legitimidad democrática. Desde esta perspectiva, los cuatro pilares de análisis: eficiencia, equidad, transparencia y legitimidad, constituyen un marco para entender como la forma en que la IA incide en las principales funciones del Estado.

Así entonces, la eficiencia es uno de los principales argumentos para la adopción de la IA en los gobiernos gracias a su capacidad de transformar recursos en resultados dado que la automatización de tareas, reducción de errores y optimización de flujos de decisión prometen incrementar la productividad estatal. Sin embargo, la eficiencia carece de neutralidad puesto que depende de cómo se definen los objetivos y de que tan equitativa sea la distribución de beneficios.

Algunos de los principales indicadores de eficiencia son: tiempo de procesamiento, reducción de costos administrativos, exactitud en la gestión de datos y capacidad de predicción. No obstante, estos beneficios deben delimitarse ya que la eficiencia no debe justificar la automatización en la toma de decisiones que afecten derechos sin mecanismos claros de supervisión humana y espacios para la apelación.

La regla operativa derivada establece que solo puede adoptarse IA en procesos críticos cuando existan métricas verificables de desempeño y salvaguardas activas que garanticen la corrección de errores.

La equidad examina cómo la IA distribuye beneficios y riesgos entre toda la población. Los modelos algorítmicos funcionan a partir de datos que, muchas veces, reflejan desigualdades estructurales.

La literatura ha demostrado que los sistemas de puntuación, clasificación o predicción suelen generar patrones de exclusión cuando los datos no son representativos, sobre todo cuando a las poblaciones vulnerables hacen referencia (Barocas & Hardt, 2020; Mehrabi et al., 2022). Los componentes que pueden explicar estos efectos pueden ser sesgos de muestreo, decisiones de diseño que priorizan la exactitud sobre la justicia distributiva y la falta de supervisión humana.

Los principales indicadores de equidad son las métricas distributivas y la falta de disparidades no justificadas. Desde una perspectiva normativa se dispone que ningún sistema automatizado debe implementarse sin una evaluación previa de impacto algorítmico ni sin mecanismos de corrección permanente.

Desde el ámbito administrativo, el uso de la IA es aceptable cuando ayuda a reducir desigualdades, o como mínimo no las incrementa y cuando se tiene salvaguardas para identificar y corregir efectos adversos.

La transparencia se refiere a la capacidad institucional de comprender cómo y por qué un sistema algorítmico produce determinados resultados. La opacidad algorítmica puede surgir por complejidad técnica, por decisiones de diseño o por restricciones contractuales impuestas por proveedores privados (Burrell, 2016). La transparencia no significa revelar el código completo, sino más asegurar que ciudadanos, supervisores y funcionarios puedan acceder a explicaciones sobre criterios, variables y razonamientos del modelo.

Los principales indicadores de transparencia son: , documentación del sistema, posibilidad de auditorías externas, trazabilidad de decisiones y divulgación de criterios generales. La regla normativa exige que toda decisión automatizada sea explicable, trazable y revisable, garantizando así a la ciudadanía el contar con vías efectivas de imanación de errores.

Finalmente, la legitimidad es el pilar más extenso ya que analiza si el uso de la IA mantiene, menoscaba o fortalece la autoridad la autoridad con base en la participación, control y rendición de cuentas. Si bien la IA puede incrementar la confianza cuando mejora servicios y coordina servicios de manera eficiente, también puede menoscabarla cuando deja de lado el juicio humano, concentra el poder de decisión en actores no electos o limita la capacidad de los ciudadanos para impugnar una decisión (Beckman et al., 2024; Yeung, 2018).

Entre los factores que afectan la legitimidad están el desplazamiento de procesos deliberativos, la excesiva delegación y la dependencia tecnológica. Por otra parte, los indicadores de legitimidad son los niveles de confianza de las personas, la claridad institucional sobre el uso de la IA y la existencia de mecanismos de apelación.

Su límite normativo es claro, la IA no puede tomar el lugar del núcleo de deliberación democrática. La regla operativa señala que los sistemas algorítmicos deben estar bajo la supervisión y aprobación del juicio humano y que las instituciones deben garantizar la transparencia, participación y control o auditoría externa.

Juntos estos cuatro ejes o pilares no como una herramienta neutral, sino como un instrumento que redistribuye riesgos, responsabilidades y capacidades.

La eficiencia sin equidad genera procesos rápidos pero injustos, la equidad sin transparencia puede producir sistemas bien intencionados pero imposibles de entender y la transparencia sin legitimidad genera sistemas claros pero antidemocráticos.

En conclusión, el marco teórico aquí propuesto integra estas dimensiones dentro de un análisis coherente que identifica tanto la evaluación de riesgos como la elaboración de defensas para una gobernanza pública responsable en contextos de digitalización acelerada.

Gobernanza algorítmica operativa

La gobernanza algorítmica en el sector público requiere un enfoque que trascienda el inventario de herramientas y se estructure como una secuencia operativa capaz de orientar decisiones institucionales antes, durante y después del despliegue de sistemas automatizados. Este enfoque reconoce que los riesgos asociados al uso de inteligencia artificial (IA) no emergen solo en la fase de implementación, sino desde la concepción del proyecto y se extienden a su operación cotidiana y a sus efectos acumulativos. Entender esta secuencia permite convertir principios éticos y normativos en procedimientos verificables con umbrales mínimos y condiciones de no implementación que no permitan lugar a dudas. La bibliografía reciente relacionada con la ética logarítmica y políticas públicas remarca que la gobernanza efectiva depende de esta estructura progresiva (Kaminski & Malgieri, 2021; Raji et al., 2020; Shneiderman, 2022).

En la fase antes de la implementación, las instituciones deben decidir si un sistema algorítmico es oportuno y justificado para el problema público que se pretende abordar. Esto involucra realizar la evaluación de impacto algorítmico que tenga en consideración los riesgos distributivos, sesgos potenciales, efectos sobre los derechos, impactos para las poblaciones vulnerables y potenciales externalidades institucionales.

Un análisis riguroso precisa considerar opciones no logarítmicas garantizando que la decisión de implementar la IA no responda a presiones tecnológicas o incentivos institucionales distorsionados.

Durante esta etapa deben establecerse umbrales mínimos que determinen cuando un sistema no debe ser implementados ya sea por ausencia de datos representativos, falta de mecanismos de supervisión

humana, definición pobre de criterios de decisión o incapacidad técnica para realizar auditorías externas. Igualmente, debe exigirse la documentación exhaustiva del sistema, incluidos objetivos, bases de datos utilizadas, supuestos del modelo y criterios de evaluación previstos (Mitchell et al., 2019).

La anterior etapa precisa que los contratos con los proveedores tengan cláusulas que garanticen la transferencia de datos, acceso a documentos técnicos, integración y depósitos de garantía sobre todo si se está hablando de sistemas de alto impacto (Zouridis et al., 2020).

Todos estos requisitos reducen el riesgo de dependencia económica y tecnológica, además de que garantizan que se seguirá operando aún si se cambia de proveedor.

En el transcurso de la implementación de la IA, la gobernanza se garantiza teniendo en funcionamiento mecanismos de control humano, seguimiento y auditorías. La supervisión humana en punto clave representa uno de los pilares normativos de la literatura contemporánea (Shneiderman, 2022). Se debe señalar que no basta con la simple presencia de un operador, el diseño del sistema debe permitir la intervención efectiva de la persona, con información suficiente para que comprenda el proceso de la toma de decisión del modelo, y debe contar también con la capacidad de cambiar, revertir o anular decisiones automatizadas.

La trazabilidad, por su parte, exige que todas las decisiones, incluidas las recomendadas por el algoritmo, queden registradas de manera que puedan ser analizadas posteriormente. Esto incluye variables procesadas, versiones del modelo, parámetros utilizados y contexto operativo. La literatura ha señalado que la ausencia de trazabilidad constituye uno de los principales factores que hacen imposible auditar los sistemas automatizados (Felzmann et al., 2020). Asimismo, la fase de implementación debe incluir auditorías externas habituales que evalúen el desempeño del modelo, la estabilidad y los efectos distributivos no anticipados. Estas auditorías deben ser realizadas por

entidades externas que no tengan conflictos de interés con el proveedor o con la agencia que implementa la tecnología (Raji et al., 2020)

También debe garantizarse la publicidad ya que los ciudadanos tienen el derecho de saber los criterios generales del sistema, su propósito, variables utilizadas y límites de su operación. La opacidad algorítmica, cuando no es técnicamente necesaria, erosiona tanto la transparencia como la legitimidad institucional (Burrell, 2016).

La fase posterior al despliegue se orienta a monitorear efectos acumulativos, corregir sesgos emergentes, evaluar impactos distributivos y garantizar mecanismos reales de apelación. Los sistemas automatizados no son estáticos: aprenden, se adaptan y, en ocasiones, se degradan con el tiempo. Es por esto por lo que la gobernanza debe considerar procesos de reentrenamiento supervisado y métodos para desactivar temporalmente cuando se manifiesten desviaciones considerables. Los impactos distributivos necesitan monitoreo fijo a través de indicadores de equidad y comparaciones entre diferentes grupos de la población, así, si se identifican efectos negativos que no pueden contrarrestarse de manera razonable, se debe activar un protocolo de suspensión del sistema.

La bibliografía también recalca la importancia de mecanismos fuertes de apelación ciudadana (Almada, 2021), que no solo implica la recepción de reclamos, sino que garantice que toda persona afectada pueda solicitar revisión humana, recibir explicaciones que pueda entender y posteriormente tener una solución efectiva a su reclamo. Además, esta etapa precisa también de evaluaciones constantes de legitimidad que analicen la percepción pública, los niveles de confianza institucional y la claridad en la comunicación gubernamental sobre el uso de la IA (Beckman et al., 2024).

En conjunto, la gobernanza algorítmica operativa se concibe como un proceso secuencial continuo que asegura que la IA no reemplace el juicio humano ni erosione principios democráticos fundamentales. Su fortaleza radica en transformar principios normativos: equidad, transparencia, participación y control democrático, en procedimientos verificables que las

instituciones puedan implementar independientemente de su nivel de sofisticación técnica. La IA tiene la capacidad para fortalecer la capacidad del Estado, pero únicamente si se implementa en un entorno que permita una continua supervisión, auditorías externas, trazabilidad completa y formas de apelación efectivas. De esta manera, lejos de ser un simple complemento técnico, la gobernanza algorítmica es la condición que determina si la IA aporta al bienestar social, o por el contrario incrementa las desigualdades y debilita la legitimidad democrática.

Casos internacionales

La evidencia internacional demuestra que la adopción de sistemas de inteligencia artificial (IA) en la administración pública produce resultados heterogéneos que dependen de los mecanismos institucionales de supervisión, la calidad de los datos y la capacidad estatal para corregir errores. Dos experiencias, una en el sector educativo y otra en la protección social, permiten observar cómo los beneficios esperados pueden verse condicionados por dilemas de equidad, transparencia y legitimidad democrática, y ofrecen lecciones operativas para orientar el uso responsable de sistemas algorítmicos.

En la educación, varios países han implementado plataformas educativas con base en la IA para personalizar los procesos de aprendizaje, identificar rezagos y prevenir deserción escolar. Un caso bastante analizado es el uso de sistemas de recomendación pedagógica en instituciones públicas de América Latina, en donde la IA se usa para ajustar contenidos, sugerir actividades diferenciadas para cada estudiante o grupo de estudiantes y monitorear el progreso de los alumnos en tiempo real (Ferreira de Lima & Veiga, 2025). El objetivo principal ha sido mejorar la eficiencia de la educación, reducir la carga laboral a los docentes y optimizar la asignación de recursos pedagógicos.

De esta manera, las primeras evaluaciones evidenciaron mejoras en el rendimiento académico de los estudiantes con acceso a infraestructura digital y en contextos institucionales donde la IA se utiliza como un complemento a

la labor del docente, más no como un sustituto. No obstante, los beneficios no se distribuyen de manera equitativa ya que las brechas de conectividad, falta de formación de los docentes y la limitada representatividad de algunos grupos en datos de entrenamiento han derivado en patrones de exclusión.

Los estudiantes de zonas rurales, con menor acceso a dispositivos digitales o conexión a internet reciben recomendaciones menos precisas o fuera de tiempo, lo que incrementa las desigualdades existentes (Capraro et al., 2024). Además el funcionamiento de estas herramientas puede llegar a ser poco comprensible para algunos docentes lo que explica por qué algunos estudiantes reciben ciertas intervenciones y otros no.

Esta falta de claridad afecta la confianza en las instituciones educativas y puede generar percepciones de arbitrariedad. En consecuencia, las conclusiones en el ámbito educativo son claras, la personalización algorítmica requiere datos representativos, supervisión docente activa, transparencia sobre criterios de recomendación y formas en las que corregir los diferentes procesos de manera que se evite que la IA incremente las brechas educativas existentes sobre todo en el contexto latinoamericano.

Una segunda área importante para el análisis de la implementación de la IA es la protección social. Varios países han adoptado sistemas automatizados para focalizar subsidios, unificar registros de beneficiarios o encontrar deficiencias en solicitudes de ayuda estatal. En Brasil el uso de sistemas algorítmicos para gestionar el Cadastro Único, una base de datos que contiene información de millones de hogares permitió mejorar la eficiencia operativa del programa Bolsa Familia, reduciendo tiempos administrativos y mejorando la coordinación entre agencias públicas (De la Briere & Lindert, 2005). La automatización permitió identificar rápidamente los hogares más vulnerables y actualizar la información para la asignación de transferencias monetarias mejorando así la eficiencia y la planificación del gasto público.

Empero, estos sistemas también han generado desafíos importantes. Por ejemplo, varias auditorías han encontrado casos de exclusión injustificada

producto de errores en la clasificación automatizada, así como dificultades para que los ciudadanos afectados impugnen estos inconvenientes.

En algunos municipios, los algoritmos operaban con datos antiguos o incompletos lo que generaba sesgos distributivos y afectaba de manera desproporcionada a la población rural personas mayores o demás población vulnerable. Las limitaciones para explicar también salieron a la luz ya que varios usuarios no entendían porque habían sido excluidos de beneficios, y en muchos casos, funcionarios tampoco podían reconstruir correctamente el razonamiento del modelo.

La experiencia brasileña resalta que la legitimidad democrática de estos sistemas no solo depende de la precisión técnica, sino también de la capacidad de los ciudadanos de entender , impugnar y corregir los fallos producto de las decisiones automatizadas.

En ambos escenarios, tanto en la educación como en la protección social la adopción de la IA en servicios públicos solo genera beneficios sostenibles cuando existe una estructura institucional que garantice la supervisión humana, auditoría externa, transparencia y métodos de corrección de errores.

La eficiencia obtenida no justifica por sí sola los riesgos que la automatización genera en términos de equidad y confianza pública. Tanto así que la evidencia internacional demuestra que los sistemas algorítmicos deben entenderse como herramientas complementarias a la intervención humana, cuyo trabajo depende del diseño institucional y del compromiso con una gobernanza democrática que garantice que toda esta innovación tecnológica se orienta hacia el bienestar público y no hacia la reproducción de desigualdades estructurales.

Discusión crítica

La integración de inteligencia artificial (IA) en la administración pública revela tensiones profundas entre los beneficios operativos que ofrece la automatización y los principios democráticos que estructuran la acción

estatal. Los ejes analíticos desarrollados: eficiencia, equidad, transparencia y legitimidad, no operan de manera aislada: se interrelacionan y, en ocasiones, entran en conflicto. La discusión crítica permite examinar estas interacciones y establecer criterios operativos para decidir cuándo un sistema algorítmico es compatible con los valores que sostienen las instituciones públicas.

En primer lugar, la eficiencia es el principal motor para la implementación de la IA. En los casos analizados queda en evidencia que la automatización mejora la velocidad de procesamiento de información, optimiza recursos e incrementa la capacidad estatal para coordinar intervenciones complicadas. No obstante, la eficiencia se torna confusa cuando no se evalúan los efectos distributivos ya que un sistema puede ser muy eficiente desde un enfoque administrativo, y a la vez generar resultados injustos y que afecten de manera desproporcionada a algunos grupos de la población. Es así como la bibliografía señala que los sistemas que utilizan la IA tienden a reforzar patrones de desigualdad cuando los datos de entrenamiento presentan sesgos históricos o cuando los mecanismos de supervisión son insuficientes (Barocas & Hardt, 2020; Mehrabi et al., 2022). En resumen, la eficiencia es admisible únicamente cuando se encuentra bajo la tutela de la equidad y no cuando la desplaza o la compromete.

En segundo lugar, la equidad se entiende como el eje que limita y orienta la utilización de la IA en procesos de alto impacto. Por ejemplo, la evidencia empírica internacional manifiesta que incluso los sistemas técnicamente bien diseñados pueden producir formas de exclusión que permanecen invisibles cuando se sustentan en bases de datos incompletas, desactualizadas o con poca representatividad.

En conclusión, la equidad requiere no solo medir los efectos distributivos, sino plantear umbrales mínimos de aceptabilidad ya que cuando un modelo genera diferencias no justificadas entre grupos, o cuando no es posible identificar las causas de dichas diferencias, deben activarse mecanismos de veto. Por lo tanto, esta regla de equidad mínima toma fuerza con la necesidad de monitoreo habitual ya que los modelos pueden desmejorarse creando aún más sesgos con el tiempo.

En tercer lugar esta la transparencia, que se traduce en la condición que permite evaluar tanto la eficiencia como la equidad ya que sin documentación técnica adecuada, trazabilidad operativa y auditorías externas, resulta imposible distinguir entre errores que son relativamente sencillos de corregir y fallos estructurales.

La discusión sobre la transparencia no se puede limitar a tener acceso al código que genera la información digital, más bien trata de la capacidad de comprender cómo funciona el sistema, cómo es el proceso de la toma de decisiones automatizadas y permitir la supervisión humana que está al tanto del funcionamiento de esta tecnología (Burrell, 2016; Shneiderman, 2022).

Los dos casos analizados demuestran que la incapacidad para explicar el funcionamiento de la IA se traduce en desconfianza por parte de la ciudadanía hacia las institucionales, y por lo tanto, la legitimidad de las decisiones es cuestionada, a pesar de que estén tomadas con base en modelos técnicos sólidos. Teniendo esto en consideración, se concluye que ningún sistema algorítmico que afecte derechos o asignaciones puede permanecer oculto, por lo tanto, la transparencia es un requisito indispensable para la corrección de errores y la rendición de cuentas.

Por último, la legitimidad democrática es el eje que une y evalúa a los demás. La IA puede fortalecer la legitimidad a través del mejoramiento de los servicios, reduciendo la arbitrariedad y ampliando la capacidad estatal. Empero, también puede debilitarla cuando reemplaza al juicio humano, limita la participación o crea una dependencia tecnológica que concentra el poder sobre todo en actores no electos (Beckman et al., 2024; Yeung, 2018). Por lo tanto, la legitimidad depende del balance entre los beneficios administrativos y los costos democráticos. Un sistema eficiente pero sin transparencia es ilegítimo y un sistema equitativo pero imposible de auditar igualmente lo es. Consecuentemente, la legitimidad requiere explicar y justificar la toma de decisiones , garantizar la apelación efectiva y asegurar que la IA complemente, más no reemplace a la deliberación pública.

La integración de los cuatro ejes permite identificar las condiciones bajo las cuales se puede implementar la IA de forma responsable y sostenida. Con base en ello se presenta una matriz que articula criterios operativos provenientes del marco teórico y de la evidencia empírica.

La matriz recomienda que la adopción de la IA es aceptable cuando los beneficios de eficiencia son verificables, medibles y constantes, no existen desigualdades entre los diferentes grupos de la población, existen fuertes mecanismos de control y evaluación y se garantiza la constante supervisión humana en cada fase de la formulación, implementación y evaluación de políticas públicas y toma de decisiones.

Por el contrario, no se debe implementar la IA si no hay datos suficientes o no son representativos, se afecta de manera negativa a la población vulnerable, los procesos no son comprensibles o no existe la posibilidad de apelación a la toma de decisiones.

En conclusión, en el sector público el valor de la IA depende de la estructura social que la acoge que de su capacidad técnica. En consecuencia, el principal desafío es integrar los beneficios de la implementación de la IA en el sector público con los valores de legitimidad del Estado.

Conclusiones y propuestas

Las transformaciones que introduce la inteligencia artificial (IA) en la administración pública no pueden comprenderse solo desde su dimensión tecnológica. Su aplicación en procesos estatales rediseña principios básicos de acción pública, modifica las relaciones entra las instituciones y los ciudadanos, y transforma los parámetros a través de los cuales evaluamos la justicia, la eficiencia y la legitimidad democrática.

Primeramente, el estudio arroja que aunque la eficiencia es un incentivo primordial para la implementación de la IA, esta no puede ser un criterio autónomo de decisión. La eficiencia administrativa es legítima solo

cuando va de la mano con la equidad distributiva y la transparencia institucional.

La automatización sin salvaguardas redistribuye los beneficios de forma injusta, más en sistemas donde los datos tienen desigualdades históricas o donde no hay suficiente supervisión humana. Es por esto por lo que la eficiencia debe regirse de manera que ningún beneficio operativo justifique la adopción de sistemas logarítmicos que generen diferencias injustificadas o reduzcan la capacidad de apelación o control ciudadano.

Por otro lado, los casos analizados muestran que incluso en sistemas técnicamente solidos se pueden generar exclusiones si los datos son insuficientes o cuando no se garantizan los mecanismos de corrección. En consecuencias, se propone que todo sistema algorítmico debe pasar por evaluaciones habituales de impacto distributivo mediante métricas que permitan evidenciar desigualdades emergentes entre grupos. La ausencia de datos representativos, herramientas de monitoreo constante o canales de revisión provocan un posible veto para su implementación.

La transparencia es la condición que tanto la eficiencia como la equidad se puedan evaluar en la práctica ya que sin trazabilidad, documentación y explicabilidad, no se puede entender el razonamiento del modelo, corregir errores o evaluar su funcionamiento. Es así como la transparencia se convierte en un principio operativo y no solo ético ya que la capacidad de auditar, explicar y revisar las decisiones automatizadas es una categoría mínima para mantener la confianza de la población.

Por último, la legitimidad democrática une y evalúa los demás ejes o pilares. La IA fortalece la legitimidad cuando mejora la calidad del servicio, reduce arbitrariedades y aumenta la consistencia de las decisiones. Sin embargo, se puede debilitar si se reemplaza el debate humano, se concentra el poder en actores no electos o se pone trabas a la apelación de decisiones automatizadas, por esto la legitimidad necesita de instituciones que garanticen la participación, supervisión constante y transparencia.

El artículo señala que la IA es factible con la gobernanza democrática únicamente cuando funciona dentro de institución que procure la equidad, transparencia y constante supervisión humana. Esta estructura institucional se evidencia en procedimientos que se puedan verificar, evaluaciones de impacto logarítmico previas, auditorías externas, mecanismos fuertes de trazabilidad y canales abiertos para la apelación. Sin estos elementos, la eficiencia es un argumento insuficiente y en muchos casos, incluso peligroso.

Finalmente, este artículo también genera aportes conceptuales ya que ofrece un marco integrado que articula mecanismos causales a través de los que la IA transforma la acción estatal, así como criterios normativos que se pueden aplicar a diferentes contextos institucionales.

En resumen, la IA ofrece una valiosa oportunidad para mejorar la capacidad del estado, sin embargo, también tiene riesgos que no se pueden ignorar. Por ende, se tiene el desafío de diseñar una estructura institucional que garantice que su implementación estará regido por los principios democráticos y que su principal objetivo es aumentar el bienestar ciudadano.

El Estado tiene la misión de implementar esta tecnología sin dejar de lado los principios y valores que dan legitimidad a las instituciones públicas. Teniendo presente que cuando se adoptan protecciones claras, estándares verificables y procesos de constante revisión, la IA se convierte en un socio estratégico para la gobernanza democrática, pero cuando estos elementos están ausentes, se puede convertir en una serie de desigualdades y desconfianza institucional.

Referencias

Almada, M. (2021). Human oversight in automated decision-making: Principles and challenges. *AI & Society*, *36*(4), 1257–1270. https://doi.org/10.1007/s00146-020-01077-3

Barocas, S., & Hardt, M. (2020). *Fairness in machine learning*. MIT Press.

Beckman, R., Müller, J., & Hansen, T. (2024). Democratic accountability in algorithmic governance. *Governance*, *37*(1), 55–74. https://doi.org/10.1111/gove.12710

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1–12. https://doi.org/10.1177/2053951715622512

Capraro, C., Hernández, L., & Paredes, M. (2024). Algorithmic personalization in Latin American public education. *Education Policy Analysis Archives*, *32*(12), 221–240. https://doi.org/10.14507/epaa.32.2411

De la Briere, B., & Lindert, K. (2005). *Reforming Brazil's Bolsa Família Program* (Social Protection Discussion Paper No. 0534). The World Bank. https://openknowledge.worldbank.org/handle/10986/13687

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2020). Transparency you can trust: Transparency requirements for artificial intelligence. *Big Data & Society*, *7*(1). https://doi.org/10.1177/2053951720921115

Ferreira de Lima, G., & Veiga, A. (2025). AI-supported pedagogical systems in Latin America: Equity and performance implications. *Journal of Learning Analytics*, *12*(1), 44–63. https://doi.org/10.18608/jla.2025.8123

Kaminski, M. E., & Malgieri, G. (2021). Algorithmic impact assessments under the GDPR. *Computer Law & Security Review*, *41*, 105532. https://doi.org/10.1016/j.clsr.2021.105532

Kroll, J. (2020). Outlining traceability: A principle for accountability in machine learning systems. *Harvard Journal of Law & Technology*, *33*(2), 403–454. https://doi.org/10.1145/3442188.3445937

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *arXiv*. https://doi.org/10.48550/arXiv.2001.00973

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, *12*(4), 505–523. https://doi.org/10.1111/rego.12158

Zouridis, S., van Eck, M., & Bovens, M. (2020). Automated decision-making in the public sector: Between efficiency and democratic legitimacy. *Information Polity*, *25*(3), 281–297. https://doi.org/10.3233/IP-200238